Software Diversification for WebAssembly

JAVIER CABRERA-ARTEAGA

Doctoral Thesis in Computer Science
Supervised by
Benoit Baudry and Martin Monperrus

Stockholm, Sweden, March 2024

TRITA-EECS-AVL-2024:10
ISBN 978-91-8040-822-6

KTH Royal Institute of Technology
School of Electrical Engineering and Computer Science
Division of Software and Computer Systems
SE-10044 Stockholm
Sweden

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av Teknologie doktorexamen i elektroteknik
.

Tryck: Universitetsservice US AB

Abstract

WebAssembly, now the fourth officially recognized web language, enables web browsers to port native applications to the Web. Furthermore, WebAssembly has evolved into an essential element for backend scenarios such as cloud and edge computing. Therefore, WebAssembly finds use in a plethora of applications, including but not limited to, web browsers, blockchain, and cloud computing. Despite the emphasis on security since its design and specification, WebAssembly remains susceptible to various forms of attacks, including memory corruption and side-channels. Furthermore, WebAssembly has been manipulated to disseminate malware, particularly in cases of browser cryptojacking.

Web page resources, including those containing WebAssembly binaries, are predominantly served from centralized data centers in the modern digital landscape. In conjunction with browser clients, thousands of edge devices operate millions of identical WebAssembly instantiations every second. This phenomenon creates a highly predictable ecosystem, wherein potential attackers can anticipate behavior either in browsers or backend nodes. Such predictability escalates the potential impact of vulnerabilities within these ecosystems, paving the way for high-impact side-channel and memory attacks. For instance, a flaw in a web browser, triggered by a defective WebAssembly program, holds the potential to affect millions of users.

This work aims to harden the security within the WebAssembly ecosystem through the introduction of Software Diversification methods and tools. Software Diversification is a strategy designed to augment the costs of exploiting vulnerabilities by making software less predictable. The predictability within ecosystems can be diminished by automatically generating different, yet functionally equivalent, program variants. These variants strengthen observable properties that are typically used to launch attacks, and in many instances, can eliminate such vulnerabilities.

This work introduces three tools: CROW, MEWE as compiler-based approaches, and WASM-MUTATE as a binary-based approach. Each tool has been specifically designed to tackle a unique facet of Software Diversification. We present empirical evidence demonstrating the potential application of our Software Diversification methods to WebAssembly programs in two distinct ways: Offensive and Defensive Software Diversification. Our research into Offensive Software Diversification in WebAssembly unveils potential paths for enhancing the detection of WebAssembly malware. On the other hand, our experiments in Defensive Software Diversification show that WebAssembly programs can be hardened against side-channel attacks, specifically the Spectre attack.

Keywords: WebAssembly, Software Diversification, Side-Channels

Sammanfattning

WebAssembly, nu det fjärde officiellt erkända webbspråket, gör det möjligt för webbläsare att portera nativa applikationer till webben. Dessutom har WebAssembly utvecklats till en väsentlig komponent för backend-scenarier såsom molntjänster och edge-tjänster. Därmed används WebAssembly i en mängd olika applikationer, däribland webbläsare, blockchain och molntjänster. Trots sitt fokus på säkerhet från dess design till dess specifikation är WebAssembly fortfarande mottagligt för olika former av attacker, såsom minneskorruption och sidokanalattacker. Dessutom har WebAssembly manipulerats för att sprida skadlig programvara, särskilt otillåten cryptobrytning i webbläsare.

Webbsideresurser, inklusive de som innehåller exekverbar WebAssembly, skickas i en modern digital kontext huvudsakligen från centraliserade datacenter. Tusentals edge-enheter, i samarbete med webbläsarklienter, kör miljontals identiska WebAssembly-instantieringar varje sekund. Detta fenomen skapar ett högst förutsägbart ekosystem, där potentiella angripare kan förutse beteenden antingen i webbläsare eller backend-noder. En sådan förutsägbarhet ökar potentialen för sårbarheter inom dessa ekosystem och öppnar dörren för sidkanal- och minnesattacker med stor påverkan. Till exempel kan en brist i en webbläsare, framkallad av ett defekt WebAssembly-program, ha potential att påverka miljontals användare.

Denna avhandling syftar till att stärka säkerheten inom WebAssembly-ekosystemet genom införandet av metoder och verktyg för mjukvarudiversifiering. Mjukvarudiversifiering är en strategi som är utformad för att öka kostnaderna för att exploatera sårbarheter genom att göra programvaran oförutsägbar. Förutsägbarheten inom ekosystem kan minskas genom att automatiskt generera olika programvaruvarianter. Dessa varianter förstärker observerbara egenskaper som vanligtvis används för att starta attacker och kan i många fall helt eliminera sådana sårbarheter.

Detta arbete introducerar tre verktyg: CROW, MEWE och WASM-MUTATE. Varje verktyg har utformats specifikt för att hantera en unik aspekt av mjukvarudiversifiering. Vi presenterar empiriska bevis som visar på potentialen för tillämpning av våra metoder för mjukvarudiversifiering av WebAssembly-program på två distinkta sätt: offensiv och defensiv mjukvarudiversifiering. Vår forskning om offensiv mjukvarudiversifiering i WebAssembly avslöjar potentiella vägar för att förbättra upptäckten av WebAssembly-malware. Å andra sidan visar våra experiment inom defensiv mjukvarudiversifiering att WebAssembly-program kan härdas mot sidokanalattacker, särskilt Spectre-attacken.

LIST OF PAPERS

1. ***WebAssembly Diversification for Malware Evasion***
   Javier Cabrera-Arteaga,Tim Toady, Martin Monperrus, Benoit Baudry
   Computers & Security, Volume 131, 2023, 17 pages
   https://www.sciencedirect.com/science/article/pii/S0167404823002067

2. ***WASM-MUTATE: Fast and Effective Binary Diversification for WebAssembly***
   Javier Cabrera-Arteaga, Nicholas Fitzgerald, Martin Monperrus, Benoit Baudry
   Submitted to Computers & Security, 2024, 20 pages
   https://www.sciencedirect.com/science/article/pii/S0167404824000324

3. ***Multi-Variant Execution at the Edge***
   Javier Cabrera-Arteaga, Pierre Laperdrix, Martin Monperrus, Benoit Baudry
   Workshop on Moving Target Defense (MTD 2022), 12 pages
   https://dl.acm.org/doi/abs/10.1145/3560828.3564007

4. ***CROW: Code Diversification for WebAssembly***
   Javier Cabrera-Arteaga, Orestis Floros, Oscar Vera-Pérez, Benoit Baudry, Martin Monperrus
   Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb 2021), 12 pages
   https://doi.org/10.14722/madweb.2021.23004

5. ***Superoptimization of WebAssembly Bytecode***
   Javier Cabrera-Arteaga, Shrinish Donde, Jian Gu, Orestis Floros, Lucas Satabin, Benoit Baudry, Martin Monperrus
   Conference Companion of the 4th International Conference on Art, Science, and Engineering of Programming (Programming 2021), MoreVMs, 4 pages
   https://doi.org/10.1145/3397537.3397567

6. ***Scalable Comparison of JavaScript V8 Bytecode Traces***
   Javier Cabrera-Arteaga, Martin Monperrus, Benoit Baudry
   11th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages (SPLASH 2019), 10 pages
   https://doi.org/10.1145/3358504.3361228

ACKNOWLEDGEMENT

First and foremost, to my beloved wife, Ilena - this journey has been yours as mine.

Five years ago, relocating to Sweden for a Ph.D. journey seemed unimaginable. At that time, I was grappling with the decision of whether to advance my academic pursuits. During this phase of uncertainty, my mentor and friend, Oscar Luis Vera, presented a life-changing opportunity: a Ph.D. position in Software Engineering at KTH. Now, as I look back over these past five years, I see a journey marked by challenges. Each step, no matter how difficult, has contributed immeasurably to my growth and learning. I carry with me a deep sense of accomplishment and no regrets—my heartfelt thanks to Oscar and Ali.

I want to thank the distinguished members of my jury. Professor Sukyoung Ryu, thanks for accepting to take the role of opponent in my defense. Professor Dilian Gurov, thanks for taking the advance review of this work. Professors, Quentin Stiévenart, Bjorn De Sutter, and Weihang Wang, thank you all for being part of my grading committee and for your invaluable insights.

I would like to extend my gratitude to my supervisors, Benoit Baudry and Martin Monperrus. Their guidance has been a constant, ensuring that I never felt lost. Their support extended beyond professional advice, blossoming into a genuine friendship. There were times when I might have been challenging to work with, but I deeply appreciate your patience and willingness to listen - all that many histories about Cuba could be boring :) - Thank you, Benoit and Martin.

Over the past five years, a significant period in my life, I have had the privilege of meeting and working with many people. Each one has played a unique role in contributing to my work. I extend my heartfelt gratitude to my lab colleagues for their unwavering support and collaboration. Special thanks to Cesar Soto, Nicolas Harrand, Romi Tsoupidi, Javier Ron, Erik Gustavsson, Nadia Campo, Orestis Floros, Long Zhang, He Ye, Deepika Tiwari, Zimin Chen, Khashayar Etemadi, Amir Ahmadian, Mikhail Shcherbakov, Yi Liu, André Silva and Anoud Alshanak for their invaluable contributions and camaraderie.

These last five years have been supported by the Stiftelsen för Strategisk Forskning(SSF) and the Trustfull project. I would like to thank all my colleagues in this project, especially Professor Musard Balliu for his invaluable insights.

To my Cuban friends in Sweden. Our roots in Cuba instill a deep sense of community and connection, often accompanied by an emotional intensity that makes each recognition meaningful. Understanding this shared sentiment, I choose to embrace all of you collectively in this paragraph without writing names. It's a way to ensure that no one feels overlooked. This is my heartfelt acknowledgment to all of you, my friends.

Last but never least, to my family in Cuba, especially to "mi abuela Mary". Family values led me to this successful journey.

Contents

Part I

Thesis